IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 8, 2020, accepted February 19, 2020, date of publication February 24, 2020, date of current version March 4, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2976134

Predicting Emerging Trends on Social Media by Modeling It as Temporal Bipartite Networks

ASIF KHAN[1], JIAN PING LI[1], NAEEM AHMAD[2], SHUCHI SETHI[3], AMIN UL HAQ[1], SAROSH H. PATEL[4], AND SABIT RAHIM[5]

[1]School of Computer Science and Engineering, University of Electronic Science and Technology of China (UESTC), Chengdu 611731, China
[2]School of Computer Applications, Madanapalle Institute of Technology and Science, Madanapalle 517325, India
[3]Department of Computer Science, Jamia Millia Islamia (A Central University), New Delhi 110025, India
[4]Department of Computer Science and Engineering, RISC Laboratory, University of Bridgeport, Bridgeport, CT 06604, USA
[5]Department of Computer Sciences, Karakoram International University, Gilgit 15100, Pakistan

Corresponding authors: Asif Khan (asifkhan@uestc.edu.cn), Jian Ping Li (jpli2222@uestc.edu.cn), and Amin Ul Haq (khan.amin50@yahoo.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61370073, in part by the National High Technology Research and Development Program of China under Grant 2007AA01Z423, and in part by the project of the Science and Technology Department of Sichuan Province.

ABSTRACT The behavior of peoples' request for a post on online social media is a stochastic process that makes post's ranking highly skewed in nature. We mean peoples interest for a post can grow/decay exponentially or linearly. Considering this nature of the evolutionary peoples' interest, this paper presents a Growth-based Popularity Predictor (GPP) model for predicting and ranking the web-contents. Three different kinds of web-based real datasets namely Movielens, Facebook-wall-post and Digg are used to evaluate the performance of the proposed model. This performance is measured based on four information-retrieval metrics Area Under receiving operating Characteristic (AUC), Novelty, Precision, and Kendal's Tau. The obtained results show that the prediction performance can be further improved if the score is mapped onto a cumulative predicted item's ranking.

INDEX TERMS Retrieval-ranking, trend prediction, recommender system, social media, information retrieval.

I. INTRODUCTION

Ubiquitous internet access is permitting users world over to be connected on social media such as Twitter, Facebook, and Youtube. This results in generation of huge volume of data every minute. It has been observed that our interaction is not limited to users but also with items. The item could be commercial-products, online-contents such as web-pages or movies. This big data can be useful in various ways, for example users liking, sharing and voting for the items on e-commerce websites can act as endorsement for other users and they may buy those products in the near future.

This data can be represented as bipartite-networks and interaction between two different kinds of nodes for user likes can be well explained. Bipartite-networks are useful in the study of market-policy of an item and personalized recommendations of an item by traversing the networks. On the other hand, monopartite - networks can help describe interaction among the same kind of nodes to represent situations where a user follows other users. This will further be helpful in finding the influential nodes and their reputation in the networks. Monitoring the behaviour of the users will provide us with prediction of items popularity or changing interest of users. As liking or sharing varies frequently social networking sites, who require the data for advertising, need to mine information continuously. Predicting future trends is of great importance in providing good marketing strategies and better use of system resources. Therefore, the analysis of such democratic and diverse data is becoming an emerging field for the research community [1]–[5].

Recently, the scope of the digital world in increased in a way that a number of smart objects are connecting in a huge quantity [6]. During the current era, a new field of research has emerged, referred to as the social IoT, which mainly includes social networking features. The social resources refers to smart devices that are capable of creating

The associate editor coordinating the review of this manuscript and approving it for publication was Guitao Cao.

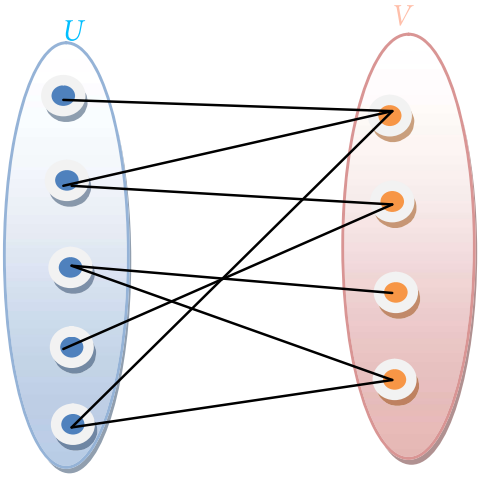

FIGURE 1. Bipartite network.

interactions with each other to independently achieve a common goal [7]. A small-world network is navigable if there exists some short route to connect all pairs of nodes in the network [8]. Due to its vast application researchers are focused on this newly derived area and genrating a huge amount of data [9]. Data is used by various online social networks, i.e. Twitter, LinkedIn and Facebook etc. analyzing and mining of useful extracted information from these social networks is not an easy task [10]. To overcome the problem of big data considering every detail of items' at every moment of sharing or liking is a difficult task, consideration to solve the problem by modeling user-item or item-item interaction over time through evolving bipartite networks and mono- bipartite networks respectively.

Bipartite networks are a particular class of complex networks, whose nodes are divided into two sets, and only connections between two nodes in different sets are allowed. Bipartite stands for an important class in social networks, and many unipartite networks can be reinterpreted as bipartite networks when edges are modeled as vertices, such as items networks. While bipartite is the special case of general graphs, common link prediction function cannot predict the edge occurrence in bipartite graph without any specialization that's why here study using the history revision information as novel popular items [11]. For the convenience of directly showing the relation structure among a particular set (U, V) of nodes as in Figure 1, bipartite networks are usually compressed by one-mode projection. This means that the ensuing network contains nodes of only either of the two sets, and two nodes are connected only if when they have at least one common neighboring node.

The huge volume of the data generated by item creation and sharing along with high speed and ease of access requires advanced analyzing methods to mine relevant information for the popularity prediction of an item. This is because very often users are attracted to popular items and the phenomena of sudden mass attraction leaves a trend on social media with an impact on social-economic and political-systems [12]. The parameter selection is another challenge. Since item-consumption is occurring at all times on the social-media, finding past-trend of item-consumption or future demand of the item is relatively easier using precise temporal details. While Predicting future popularity based on the item feature is a difficult task [13]. This is especially true, when all features of items cannot be considered at run-time, because every instant online content is created without explicit user features such as Facebook, Youtube, and Twitter content.

Based on the extensive review of previous works, important research opportunity for the presentation has been identified. Mainly one can find lack of models for (1) novel item predictions and (2) correct prediction of temporal popular items.

According to overcome the above limitations, these problems have been solved by including many aspect and system into account. Considering recent popularity, three different models depending on dominant factor, non-dominant factor and aging effect have been proposed to predict long term and short term treads. Among them recent popularity dominant model performs better than the rest. Also it is worth mentioning than when parametric recent popularity has been considered along with the aging effect, it help to discover the temporal popular items too.

In this paper, we propose a model to predict and rank the web-content of online social media. Using this model, we study three different kinds of online-social media namely Movielens, Facebook- wall-post and Digg. This study focuses on predicting futuristic popularity of an item based on information at a given point of time along with growth ranking which will generate trend as to which item may get maximum user attention. Progression with help of considerable empirical observations on the chosen datasets enable us to support obtained results. The said datasets contain the latest information provided by the research group about user to user and user with item interactions which are assumed as monopartite and bipartite networks respectively. We then model these networks to predict the items popularity.

Our contributions summarized as follows:

- We propose a model to predict and rank the web-content of online social media. We applied on three real datasets namely Movielens, Facebook-wall-post and Digg.
- This study focuses on a given time point along with growth ranking and makes an effort to predict which item may get great attention of the users in the given future-time-window.
- Proposed model makes prediciton of already popular items as well as it also able to predict some newly popular items.
- We provide considerable empirical observations on the chosen datasets to support obtained results. The said datasets contain the latest information provided by the research group about user-user and user-item interactions which are assumed as monopartite and bipartite networks respectively. We then model these networks to predict the items' popularity information.

The remainder of this paper is organized as follows. In Section II, we outline the related works and the motivation for our work. The proposed Growth-Based Popularity-Prediction model and its Evaluation Metrics are introduced in Section III. In Section IV, the remarkable performance is demonstrated by the Experimental Setup and Results. In Section V, Results and Discussion included. Finally, in Section VI, we summarize our proposed work.

II. RELATED WORKS

In recent years, research is intensifying in Machine Learning driven applications such as robot vision [14], [15],autonomous vehicles [16], sibilance security [17],. The medical data analysis using machine learning techniques are more suitable for diagnosis of critical disease [18]. Application-level semantics of streaming video sources are becoming more and more ubiquitous in a wide spectrum of applications. Images [19], videos and audio can provide rich data sources, from which additional information and context can be surmised.

Predicting and ranking of web-content can provide valuable insights in varied areas such as online marketing, good marketing-strategies, social-economical trend prediction and so on. As a visual asset with diverse applications, it gets focused attention of the researchers to work in this domain [18] along with addition of novelÂ prediction methods for different types of web content in recent years. Most of the contemporary literature delibrates on predicting various real-life outcomes like item recommendations, election results, social-bot identification, box-office revenues and many more using online social media [16]. Prediction techniques measure a certain level of interest for a web-content that the online-community will exihibit in near future. In the initial phase, researchers ascertained the web-access pattern and substantiated that the behavior of users requests for web-content is a stochastic process. Further, Zipfs law described that the distribution of users requests for web-content is highly skewed [22].

The web-content access pattern becomes a parameter in prediction of web- content popularity. Time series model demonstrates the measures of interest by defining rise and fall pattern of the web-content in the recent-time-window [15], [16]. For instance ephemeral fashion, movies, election-campaign exhibit continuously changing popularity trend pattern. So does online news articles. On the other hand, some events on social-media are seasonal or cyclic such as election-period and clothing. Researchers have found that some web-content gains popularity at regular intervals and study how time distribution helps in predicting cascade recurrence [6]. This is the case with bibliographic data where paper-citation often remain popular for long lifespan and gradually their popularity may be lost. However, paper-citation does sometimes exhibit sudden rise in popularity after a long hibernation period. Some studies have revealed that the interest generated by a web content is transient, heterogeneous, and often unpredictable [10], [18] if the patterns vary sharply.

Some researchers concluded that accurate prediction of web-content is very hard without early web access information [3], [19]. To tackle the issue, researchers classify the web- content popularity into two classes: Reference-based popularity prediction and View-based popularity prediction. Reference-based popularity prediction has found applications in news-articles, Facebook-wall-post and trending images/videos. Hashtag on social media characterises these web-content to predict their popularity. A proportion of literature refers to these methods as Pre-publication prediction methods. On the other hand View-based prediction methods predict web-content popularity based on users request received after their publication. These methods also called as Post-publication prediction methods and are used for bibliographic content [20].

Most often access patterns of web-contents exhibit similar popularity initially and gives a diverse pattern afterwards. These diverse popularity evolution patterns are further investigated in [4], [23] and the authors have concluded that the use of recent popularity and total popularity in a given point of time leads to higher accuracy in predictions in targeted future-time-window. To substantiate the above assertion, Zeng *et al.* [23] divided three different datasets into two time frames namely past-time-window and future- time window, and predicted that recent popularity is useful for short lifespan window and total popularity performs reasonably well in the long run.

Another model is proposed [12] in which re-tweet pattern is analyzed. This is obtained by having the tweet count for a short span of time and then predicting the re-tweet count for three days after publication. Authors first gathered data for the re-tweet count after every 20 minute time interval for first hour and then used this information to determine the most similar tweet pattern in the training set of the given re-tweet features set. The predicted popularity was then set to the weighted average of the re-tweets. Search engines predict the social media item popularity on the basis of their usage and click rate. The ranking is then obtained by predicting the future consumption of items. Although various ranking methodologies can be applied, our focus is temporal features like time of item consumption, or current trending items on social media such as popular videos on Youtube [23].

A model is proposed [12] in which re-tweet pattern is analyzed. This is obtained by having the tweet count for a short span of time and then predicting the re-tweet count for three days after publication. Authors first gathered data for the re-tweet count after every 20 minute time interval for first hour and then used this information to determine the most similar tweet pattern in the training set of the given re-tweet features set. The predicted popularity was then set to the weighted average of the re-tweets.

Item popularity is affected by mainly four categories of features namely structure, content, early adoption and temporal features. As per some researchers content is useful [A] while as per some [A, B] it is not. Most often access patterns of web-contents exhibit similar popularity initially and gives a

diverse pattern afterwards. These diverse popularity evolution patterns are further investigated in [4], [23] and the authors have concluded that the use of recent popularity and total popularity in a given point of time leads to higher accuracy in predictions in targeted future-time-window. To substantiate the above assertion, Zeng *et al.* [23] divided three different datasets into two time frames namely past-time-window and future- time window, and predicted that recent popularity is useful for short lifespan window and total popularity performs reasonably well in the long run.

Search engines predict the social media item popularity on the basis of their usage and click rate. The ranking is then obtained by predicting the future consumption of items. Although various ranking methodologies can be applied, our focus is temporal features like time of item consumption, or current trending items on social media such as popular videos on Youtube [23]. Temporal features have been given preferred treatment over other features and dominate active research. Popularity of temporal features is fortified with the fact that they are easily available and also they are independent of the item level features saving cost of feature engineering. Thus models based on these find varied applications such as Twitter, citation count and so on.

As we find the nature of web-content Evolutionary, we proposes a Growth-based Popularity Predictor (GPP) model in this paper for predicting and ranking the web-contents. To evaluate the performance of the proposed model, we have employed three actual web-based datasets namely Movielens, Facebook-wall-post and Digg. The performance of the model is measured based on four information-retrieval metrics, which are Area Under receiving operating Characteristic (AUC), Novelty, Precision and Kendals Tau. The obtained results show that the prediction performance can be further improved if score is mapped onto cumulative predicted-items ranking.

An exhausted reviews of above contributions has been carried out which motivates us to address evolutionary scenario of web-contents on online media. The novelty of the proposed approach lies in its model that works to predict novel items and correct prediction of temporal popular items. Unlike previous approaches, scores of items are assigned based on its cumulative values defined by function which depends on threshold value. This parameter makes it possible to tune the contribution of the cumulative function.

III. GROWTH-BASED POPULARITY-PREDICTION

In the era of online-social-media (OSM), millions of data are generating at each minute. Analyzing the details of items' sharing or liking of every minute is a challenging task. An analyzing method eases the information-retrieval from such evolutionary scenarios. Considering this idea of analysis, we have adopted the evolutionary scenario of a user-item graph of Facebook-wall-post along with two other datasets Movielens and Digg. In figure 2, a snapshot of Facebook-wall-post interaction user-item graph is plotted which shows how the interaction network is growing after the publication

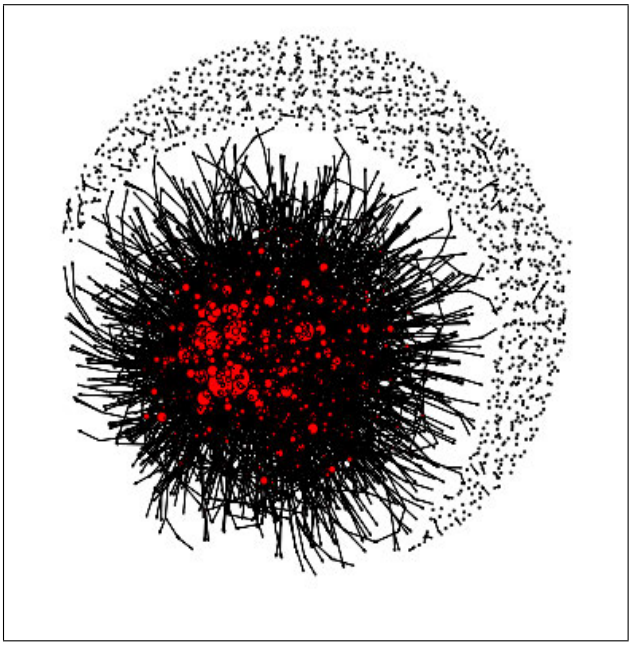

FIGURE 2. A snapshot of Facebook wall-post interaction where node-size shows likes, comments etc.

of posts. To ease this evolutionary scenario, we construct the subset of the data in the recent time-window where likes, rating, re-shares, and voting is considered as popularity. We described all these datasets in greater detail in Section IV.

Preferential Attachment-based model (PA) [20], also called rich-get-richer, best describes the network evolution assuming that *higher the degree higher the probability of attracting new links*. In our context we mean that higher popularity of an item, higher the probability of attracting new users. In this paper, authors' main aim is to predict the expected items' popularity that may be of great interest of the peoples in near future. To predict the expected popularity, a live experiment is carried out in Facebook-wall-post, Movielens and Digg. To this end, 100th most popular items are considered in the past and future time window in our experimental setup. Whenever a user's request is received for an item from the network, it validates the request and creates a map of node and list of item as link received. Since these datasets are very large and unstructured, all datasets are mapped into adjacency matrix A whose elements are 0 and 1. This is time-dependency matrix which is used to evaluate item-degree $Z_i(t) = \sum_u A_u i(t)$ indicating the number of users u who collected item i.

Unlike other social-media, we have complete information consisting of like, sharing and comments along with complete users's account-information in Facebook. This collective information is considered as review rating which is then mapped to our binary data by applying item with any of 3 rating is marked as collected by a respective user. In Movielens, any item rated 3 or above is marked as 1 in the mapped data and item is marked as collected by user if respected user voted that item. The nature of the collected data is shown in figure 4.

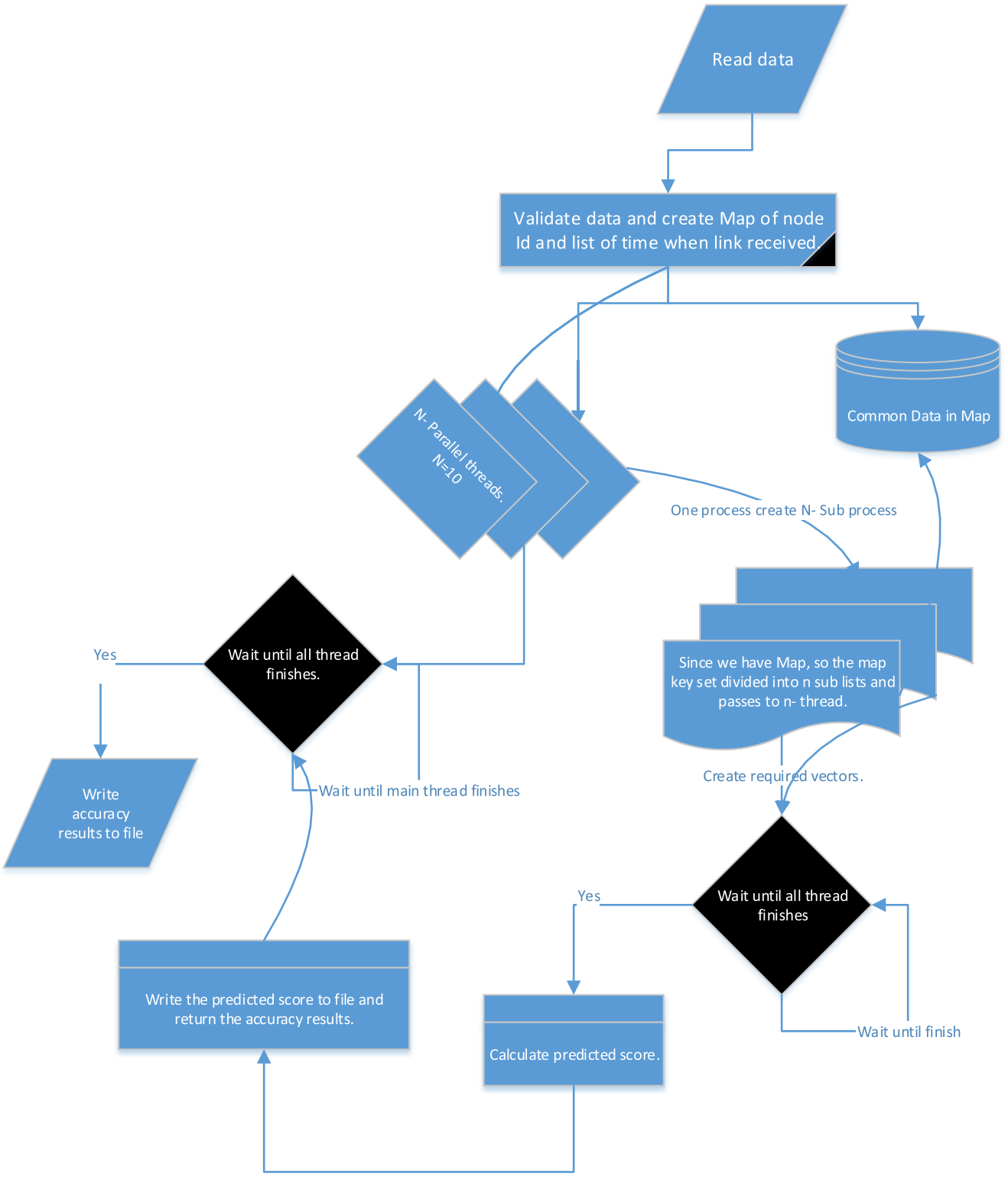

FIGURE 3. Conceptual Diagram.

As the model proposed by Lerman and Hogg described the temporal evolution of items' popularity as stochastic process of users' surfing on social-media [21]. And also in such cases, users' request of sharing or liking is highly skewed [22]. Therefore, items' rating in this paper is normalised to maintain the trade-off between consequences of too small and

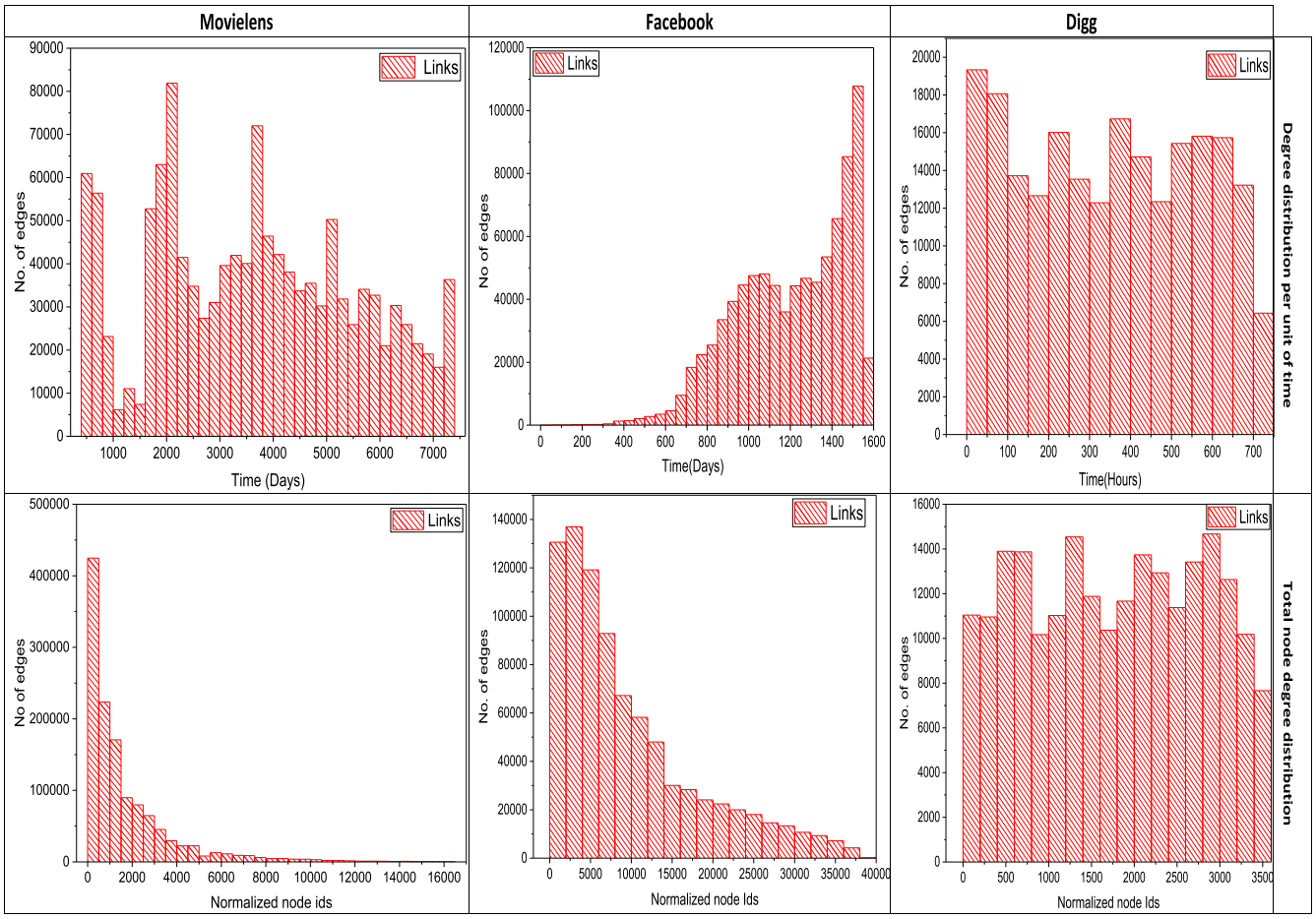

FIGURE 4. Data Statistics Plots.

too high value of T_P as mentioned in [23]. Note that $Z_i(t)$ is the normalised-stale-popularity we mean items' degree received during recent-time-window T_P. When we say normalised-current-popularity (i.e. $\Delta Z_i(t)$) we mean the items' degree increased at the time t. On the basis of these two complementary sources of information, we have introduced a discrete distribution function (Eq. 1) assigning possible score S_i labelled by compared-popularity.

$$S_i(t, T_P) = \begin{cases} T(Z_i), & If\ \Delta Z_i > Z_i \\ T(\Delta Z_i), & If\ \Delta Z_i \leq Z_i \end{cases} where, \qquad (1)$$

Scores of items are assigned based on its cumulative values defined by function $F_X(x)$ which depends on threshold value $k \in [0, 1]$. This parameter k makes it possible to tune the contribution of the cumulative function. We assumed if the item's current-popularity is greater than that of recent-popularity in stale-time-window, then scores grows exponentially for such items, otherwise less popular items's score grows linearly as shown in Eq. 2. Therefore, we refer to this model as Grows-based Popularity Predictor (GPP). The popularity growth function is defined as:

$$T(x) = \begin{cases} e^x, & If\ F_X(x) > k \\ x, & Otherwise. \end{cases} \qquad (2)$$

In our approach, we define a test date t^* and future-time-window of length T_F. The increased popularity of item i in the defined window is introduced in Eq. 3.

$$\Delta Z_i(t^* + T_F, T_F) = Z_i(t^* + T_F) - Z_i(t^* + T_F - T_F) \qquad (3)$$

All items are ranked according to their increased popularity $\Delta Z_i(t^* + T_F, T_F)$. This ranking is referred as *true ranking*. We then assign scores to all items using generic predictor as in Eq. 1 which are mapped into a *predicted ranking* to compute the expected popularity prediction. To test the performance proposed predictor, estimated ranking of top n items is computed and then are examined in order to determine their presence in the true ranking as shown in table 1. Some new items may get ranks in the top n places of estimated ranking (e.g. items i_5 and i_7 in the given table).

The number of items in the top n places of estimated ranking, also appear in the top n places of true ranking, is measured in precision $P_n \in [0, 1]$ (higher the precision the better prediction). This precision is evaluated on average

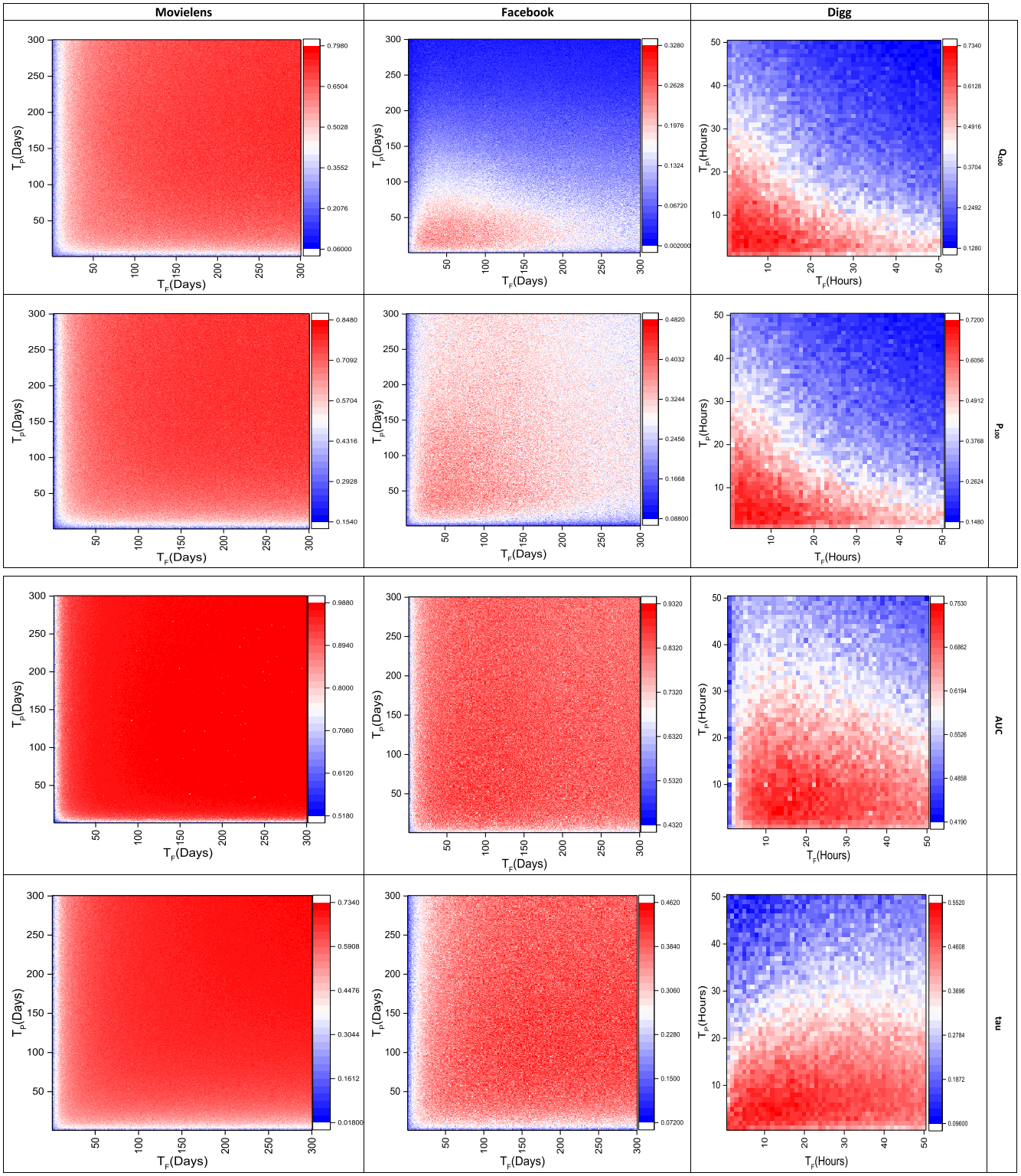

FIGURE 5. **Heat map of precision 100 in the past/future time window plane(T_P, T_F).**

results over 10,10 and 7 regularly spaced test-dates to test the performance of the proposed predictor for Movielens, Facebook, and Digg respectively. On the other hand, the number of "new entries" (got the place in the estimated ranking that missed the place in the true ranking) is measured in quality Q_n which is the ratio of number of new entries E_n to the number of items identified by predictor C_n. This information retrieval metric Q_n (i.e. E_n/C_n) shows how well the proposed

TABLE 1. Ranking table.

Items	True Ranking	Estimated Ranking	Remark
i_1	r_1	r_1^*	Identified
i_2	r_2	r_2^*	Identified
i_3	r_3	....	
i_4	r_4	r_4^*	Identified
i_5	....	r_5^*	New Entry
i_6	r_5	....	
i_7	....	r_7^*	New Entry
..	..	..	..

model is able to predict the future trends for unseen events. In our analysis, top 100 items are considered to test the performance of the model. Further details of the metrics is given in section IV.

A. EVALUATION METRICS

To measure the performance of our model we selected four information retrieval based metrics:*precision*(P_k), *novelty*(Q_k) and *Area Under receiving operating Characteristic* (AUC_k) also known as ROC [24], and Kendal's rank correlation Tau(τ).

- *Precision* is defined as the fraction of objects listed in the top k rankings of the predicted and real ranking lists [25] and is given by:

$$P_k = \frac{D_k}{k}, \tag{4}$$

where D_k is the number of common objects in the top k of both predicted and real ranking lists. $P_k \in [0, 1]$. The higher value of P_k, the better precision of prediction.

- *Novelty(Q_k)* measures the ability of a predictor to rank 'new objects' in the top k position which were not in the top k positions in the past. Let R_k denotes the number of 'new objects' in the top k position of the real list and E_K denotes the number of new object correctly predicted by our model in top k ranking list. Then the novelty score is measured by-

$$Q_k = \frac{E_k}{R_k}, \tag{5}$$

- *AUC* measures the importance of the relative position of its top k objectives in the predicted and ranked lists. It selects the top k objects from the real list as a benchmark and compares their rank scores with the top k objects in the predicted list. Let $s_p \in L_p$ and $s_r \in L_r$ be the scores of an object in predicted list. Then *AUC* can be calculated by:-

$$AUC = \frac{\sum_{s_p \in L_p} \sum_{s_r \in L_r} I(s_p, s_r)}{|L_p|\,|L_r|} \; where, \tag{6}$$

$$I(s_p, s_r) = \begin{cases} 0, & \text{if } s_p > s_r, \\ 0.5, & \text{if } s_p = s_r, \\ 1, & \text{if } s_p < s_r. \end{cases} \tag{7}$$

- *Kendal's Tau(τ)* is used to measure rank correlation between two list, i.e predicted and real ratings. It varies between -1 and $+1$. $\tau = 1$ when predicted and real (actual) are identical, $\tau = 0$ when both ranking are independent and $\tau = -1$ shows they perfectly disagree. It can be given as-

$$\tau = \frac{C - D}{\sqrt{(C + D - N_{tp})}\sqrt{(C + D - N_{tr})}}, \tag{8}$$

where C is the number of concordant pairs and D is the number of discordant pairs. N_{tp} is the number of ties in predicted list and N_{tr} number of ties in real list.

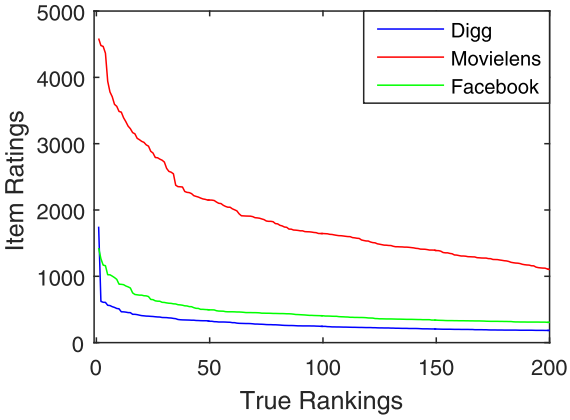

FIGURE 6. Rankings of the top 200 items in the given datasets.

IV. EXPERIMENTAL SETUP AND RESULTS

This paper presents details of data collection process, information retrieval metrics, relative comparisons of methods and their results in details Figure 3 in the following subsections. Where conceptual diagram represents the preprocessing of the data used in our analysis. Basically conceptual diagram is visual representation of the way in which abstract concepts are related. Conceptual diagrams are widely employed in fields as diverse data be created in a variety of ways to suit the items process's needs. After reading data, by taking an average of the validations and create map of node, Id and list of time when link received. Where parallel threads wait until all thread finishes. Write accuracy results to file then write the predicted score to file and return the accuracy results. Otherwise after reading common Data in Map, since we have Map, so the map key set divided into sub lists and passes to threads. Till that wait until all thread, after all finishes calculate predicted score.

A. DATA COLLECTION

In this study, we used data from Movielens, Facebook-wall-post, and Digg. Movielens is web-based movie recommendation service, currently focusing on predicting popularity of movies. Initially, this website is created by GroupLens Research gather the research data to solve the research problems. In our analysis, we used latest dataset[1]of Movielens which contains 26, 024289 rating records of 45, 843 movies

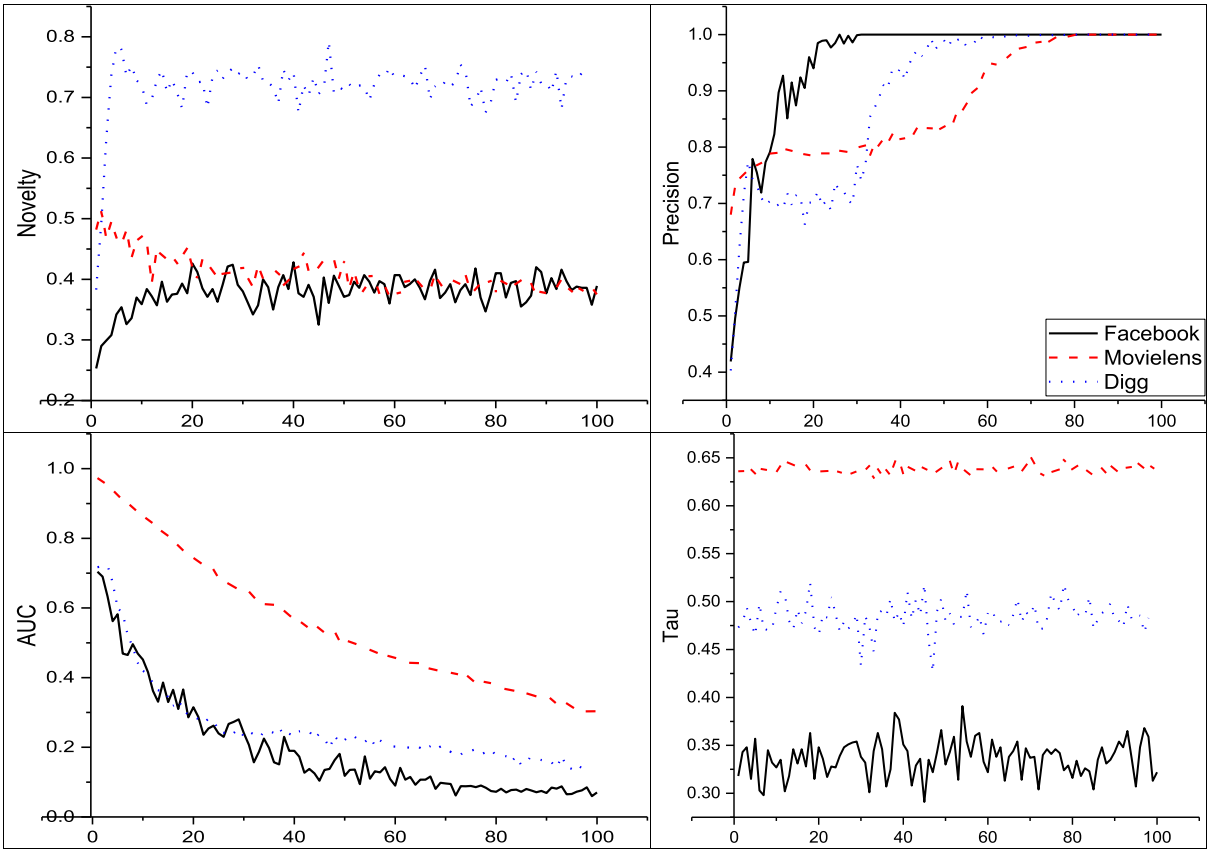

FIGURE 7. Sensitivity of model for varying top k value (number of selected objects for comparing predicted and real ranking lists), for fixed T_P and T_F. Here past and future time window i.e T_P & T_F is fixed as 30 days (Facebook and Movielens) and 10 hours for Digg dataset. The X-axis is for top k node list size as a percentage of the whole list.

rated by 270, 896 users. This data is recorded between January 09, 1995 and August 04, 2017. Each user has rated a movie from 1 to 5. In this study, we considered only positive ratings (higher than 2). To avoid biasedness we randomly selected 20 millions unique users and all the movies rated by them. We have considered the time in days. Similar to Movielens, Digg includes popular news with highly popular stories. Digg[2]data contains 3553 news items where 139409 users have voted or Digged (3018196 links) for the news stories from 31 January 2009 to 5 July 2009. We have randomly sampled 10, 000 users and all their digged news. In addition Facebook is one of the most popular social networking websites where users can share their moments, photos, videos, newsfeeds and their thoughts. Facebook[3]contains Facebook user's wall post activities from 14 October 2004 to 21st January 2009. It contains 46951 users and their wall post activity [2], [26], [27]. We remove the records where users have posted on their own wall.

All these datasets are divided into 2 parts according to 2 time-windows. First part of the dataset lie in the past-time-window and second part of the dataset lies in the future-time-window. Middle part (common in both time-windows) of the dataset is chosen for random selection of the test date t^* which is averaged over 10 regularly spaced test date for all 3 datasets. This is because we can have space for the random test-date in both time-windows. We can have enough information from the historical data, if test-date is randomly selected from past-time-window. On the other hand, we can have enough information from the data of future-time-window, which is not yet obvious, if test-date is randomly selected from the future-time-window.

In data processing, we observed that few items received very high number of ranking, while majority of items only get low ranking. All these items are sorted by their decreasing ranking. Figure 6 exhibits that the observed behaviour of many least popular items is linear as chosen linear function in Eq. 2.

V. RESULTS AND DISCUSSION

To perform the experiments we randomly selected the time-window assuming we have enough historical information of item's rating in given time-frame. We have ranked the items according to its score and then compare items' ranking according to true ratings after future time window T_F. To avoid the experimental biasness we took an average of 10 experiments considering the 10-cross fold validation in

[1]www.grouplens.org/node/73
[2]www.isi.edu/integration/people/lerman/downloads.html
[3]http://konect.uni-koblenz.de/networks/facebook-wosn-wall

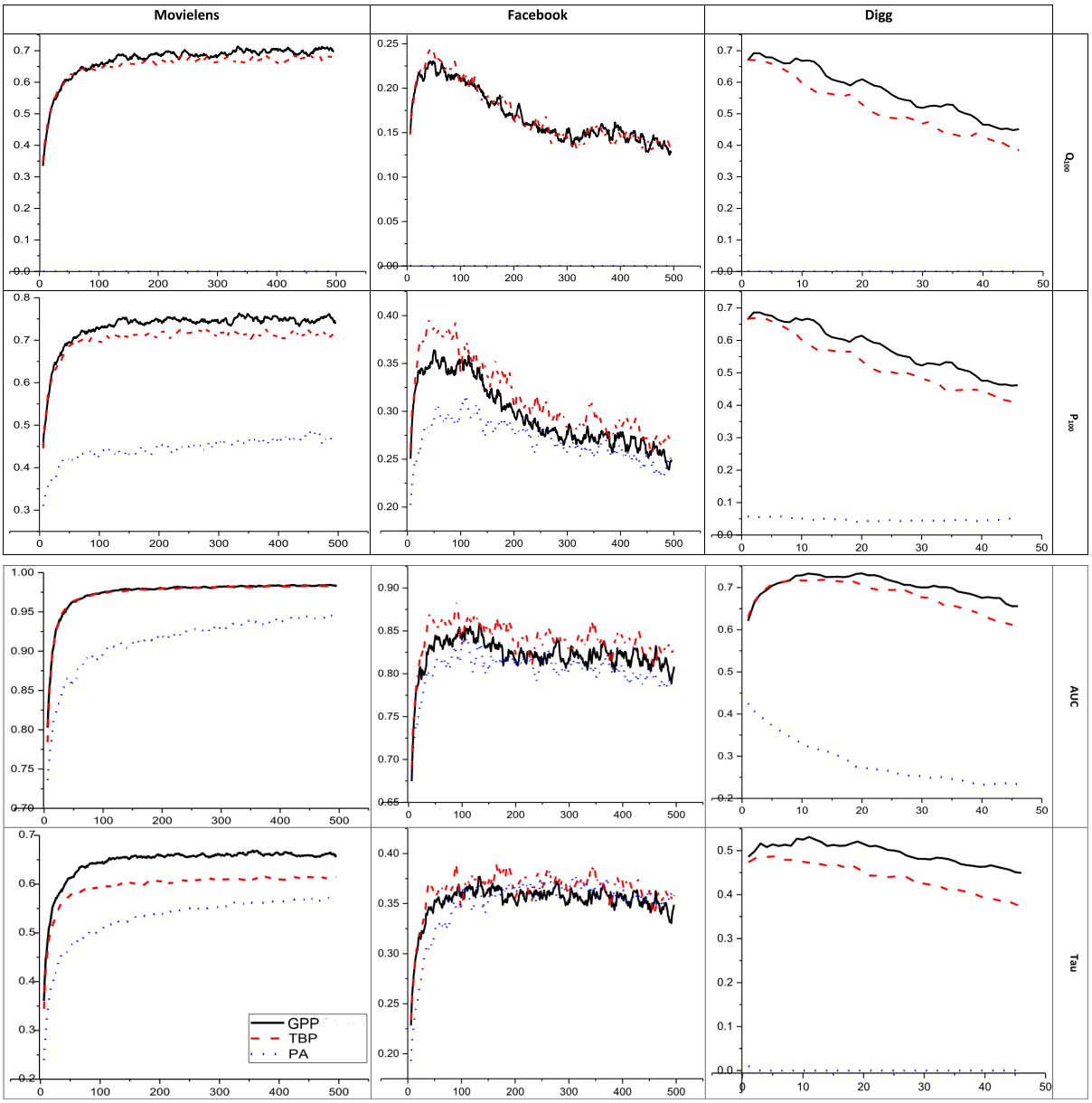

FIGURE 8. Performance of our model for fixed past and varying future time window. Here The X-axis is for time and The Y-axis is for accuracy score.

which snapshots of said networks are considered at 10 different time-steps using 10 different threads. In this section, we examine the sensitivity of our proposed model to the variable k, based on the percentage of the selected list compared to the full list, while fixing the T_P and T_F. We selected the past and future time windows as 30 days for Facebook and Movielens and 10 hours for Digg.

We examine the sensitivity of our proposed model to the variable k. As shown in Fig. 7, Precision P_k increases with the same rate in all the data sets. Novelty prediction (Q_k) becomes fixed after a threshold ($\approx$ 20), meaning it is not affected by the size of the list or in other words, is not sensitive to the k. The performance of tau (τ) doesn't get affected by the list size because it works on the whole list, not on top k items. We can see AUC decreases as the size of the list increases. In addition, Fig. 7 shows the precision of our proposed model is perfect after a threshold of about 35%. Therefore, this is an accurate model when having high k value is not computationally costly [23].

A. VARYING FUTURE TIME (T_F), FIXED PAST TIME (T_P) AND FIXED SIZE K

To set an appropriate fixed past time window (T_P) characteristics of the dataset such as their evolution rate has been considered. As the movie rating process is slower than the Facebook wall communication or Digg item prorogation. We set a longer period of 90 days as T_P for Movielens but

30 days and 5 hours for Facebook and Digg respectively. We have tested the predictor for the varying future time lengths from day 1 to 500 days for Movielens and Facebook; and up to 50 hours for Digg. The X-axis shows the time and Y-axis shows the accuracy results based on different evaluation metrics. Our proposed Model (given in Eq. 1) outperforms the PA model in all the cases except in Movielens and Facebook rank correlation case as shown in Fig. 8.

To overcome the problem of big data considering every detail of items at every moment of sharing or liking is a difficult task, this study only considers recent time windows. So, assumptions about datasets nature like keep changing (as in case of Diggs), this study consider the more complex scenario and therefore considering three time windows for already papule novel items on behalf of other baseline experimental outcomes of already popular items that has been captured in past, just before now.

VI. CONCLUSION

This paper studied the popularity prediction method of web-items available on the online-social-media. An expressive model has been proposed to predict such items' popularity that remains attractive or less attractive for longer periods of time. The model includes a more important steps with the condition that the evolution of web-content over time grow exponentially or sometimes remain linear. To measure the performance of model, 3 different real data sets and preferential attachment-based model as a benchmark have been used. The proposed model outperforms the benchmark model on these real data-sets. We have considered the data sets to keep a different kind of evolution in mind, such as on Movielens the evolution is slower than Facebook and Digg. While Digg items' evolution is faster since news items don't last for a long time. In all the cases we have found recent popularity based our proposed model outperforms. We have found that performance of the model gets better in the system where items evolve faster such as in case of Digg. We can find that our model outperforms with great margin as we move from slower system to faster system such as Movielens to Digg. In case of Movielens, our model outperforms but not with great margin as in case of Digg. In our analysis of Movielens data, we have found that recent popularity is more correlated to actual popularity gain as compared to total popularity which supports the preferential attachment.

AUTHOR CONTRIBUTIONS

Authors individual contributions in research articles are conceptualization,methodology and writing–original draft prepared by Asif Khan, Naeem Ahmad and Shuchi Sethi. Worked on Writing–original draft preparation. Validation, formal analysis, and supervision done by Jian Ping Li and Sarosh H Patel. Amin Ul Haq work on facts visualization.

CONFLICTS OF INTEREST

The authors are declaring that there is no conflict of interest.

The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

The authors confirm that the data supporting the findings of this study are available within the article.

ACKNOWLEDGMENT

All resources used for experiments support given by 'Key Laboratory of Wavelet Active Media Technology'. Room No: B1301, School of Computer Science University of Electronic Science and Technology of China (UESTC), No. 2006, Xiyuan Avenue, West Hi-Tech Zone, Chengdu, Sichuan, 611731, P. R. China.

REFERENCES

[1] K. Abbas, L. Xin, and S. Mingsheng, "Discovering items with potential popularity on social media," in *Proc. IEEE 14th Intl Conf Dependable, Autonomic Secure Comput. (DASC)*, Aug. 2016, pp. 459–466.

[2] K. Abbas, M. Shang, X. Luo, and A. Abbasi, "Emerging trends in evolving networks: Recent behaviour dominant and non-dominant model," *Phys. A, Stat. Mech. Appl.*, vol. 484, pp. 506–515, Oct. 2017.

[3] G. Szabo and B. A. Huberman, "Predicting the popularity of Online content," *Commun. ACM*, vol. 53, no. 8, pp. 80–88, Aug. 2010.

[4] E. Ferrara, R. Interdonato, and A. Tagarelli, "Online popularity and topical interests through the lens of instagram," in *Proc. 25th ACM Conf. Hypertext Social Media*, 2014, pp. 24–34.

[5] P. Bao, H.-W. Shen, X. Jin, and X.-Q. Cheng, "Modeling and predicting popularity dynamics of Microblogs using self-excited Hawkes processes," in *Proc. 24th Int. Conf. World Wide Web (WWW)*, 2015, pp. 9–10.

[6] F. Amin, A. Ahmad, and G.-S. Choi, "To study and analyse human behaviours on social networks," in *Proc. 4th Annu. Int. Conf. Netw. Inf. Syst. Comput. (ICNISC)*, Apr. 2018, pp. 233–236.

[7] F. Amin, A. Ahmad, and G. Sang Choi, "Towards trust and friendliness approaches in the social Internet of Things," *Appl. Sci.*, vol. 9, no. 1, p. 166, Jan. 2019.

[8] F. Amin, R. Abbasi, A. Rehman, and G. S. Choi, "An advanced algorithm for higher network navigation in social Internet of Things using small-world networks," *Sensors*, vol. 19, no. 9, p. 2007, Apr. 2019.

[9] F. Amin and M. Zubair, "Energy-efficient clustering scheme for multihop wireless sensor network (ECMS)," in *Proc. 17th IEEE Int. Multi Topic Conf.*, Dec. 2014, pp. 131–136.

[10] F. Amin, A. Ahmad, and G. S. Choi, "Community detection and mining using complex networks tools in social Internet of Things," in *Proc. IEEE Region Conf.*, Oct. 2018, pp. 2086–2091.

[11] Y.-J. Chang and H.-Y. Kao, "Link prediction in a bipartite network using wikipedia revision information," in *Proc. Conf. Technol. Appl. Artif. Intell.*, Nov. 2012, pp. 50–55.

[12] M. Tsagkias, W. Weerkamp, and M. De Rijke, "Predicting the volume of comments on online news stories," in *Proc. 18th ACM Conf. Inf. Knowl. Manage.*, 2009, pp. 1765–1768.

[13] C. S. Lee and L. Ma, "News sharing in social media: The effect of gratifications and prior experience," *Comput. Hum. Behav.*, vol. 28, no. 2, pp. 331–339, Mar. 2012.

[14] A. Khan, J.-P. Li, A. Malik, and M. Y. Khan, "Vision-based inceptive integration for robotic control," in *Soft Computing and Signal Processing*. Singapore: Springer, 2019, pp. 95–105. [Online]. Available: https://link.springer.com/chapter/10.1007/978-981-13-3393-4_11

[15] A. Khan, J.-P. Li, and R. Ahmed Shaikh, "Content based object observation for image retrieval," *Int. J. Comput. Appl.*, vol. 113, no. 5, pp. 18–21, Mar. 2015.

[16] A. Khan, J.-P. Li, R. A. Shaikh, and I. Khan, "Vision based classification of fresh fruits using fuzzy logic," in *Proc. 3rd Int. Conf. Comput. Sustain. Global Develop. (INDIACom)*, Mar. 2016, pp. 3932–3936.

[17] A. Khan, J.-P. Li, and R. Ahmed, "Vision based geo navigation information retreival," *Int. J. Adv. Comput. Sci. Appl.*, vol. 7, no. 1, pp. 397–401, 2016.

[18] A. U. Haq, J. P. Li, M. H. Memon, J. Khan, A. Malik, T. Ahmad, A. Ali, S. Nazir, I. Ahad, and M. Shahid, "Feature selection based on L1-norm support vector machine and effective recognition system for Parkinson's disease using voice recordings," *IEEE Access*, vol. 7, pp. 37718–37734, 2019.

[19] A. Khan, S. Deep, J.-P. Li, M. H. Memon, R. A. Shaikh, and K. Kumar, "Inchoative integration of content based image retrieval: Shodhani," in *Proc. 11th Int. Comput. Conf. Wavelet Actiev Media Technol. Inf. Process. (ICCWAMTIP)*, Dec. 2014, pp. 289–292.
[20] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, Oct. 1999.
[21] K. Lerman and T. Hogg, "Using a model of social dynamics to predict popularity of news," in *Proc. 19th Int. Conf. World Wide Web (WWW)*, New York, NY, USA, 2010, pp. 621–630.
[22] A. Tatar, M. D. de Amorim, S. Fdida, and P. Antoniadis, "A survey on predicting the popularity of Web content," *J. Internet Services Appl.*, vol. 5, no. 1, Aug. 2014.
[23] A. Zeng, S. Gualdi, M. Medo, and Y.-C. Zhang, "Trend prediction in temporal bipartite networks: The case of Movielens, Netflix, and Digg," *Adv. Complex Syst.*, vol. 16, no. 04n05, Oct. 2013, Art. no. 1350024.
[24] J. A. Hanley and B. J. McNeil, "The meaning and use of the area under a receiver operating characteristic (ROC) curve," *Radiology*, vol. 143, no. 1, pp. 29–36, Apr. 1982.
[25] J. L. Herlocker, J. A. Konstan, L. G. Terveen, and J. T. Riedl, "Evaluating collaborative filtering recommender systems," *ACM Trans. Inf. Syst. (TOIS)*, vol. 22, no. 1, pp. 5–53, Jan. 2004.
[26] B. Viswanath, A. Mislove, M. Cha, and K. P. Gummadi, "On the evolution of user interaction in facebook," in *Proc. 2nd ACM Workshop Online Social Netw. (WOSN)*, 2009, pp. 37–42.
[27] A. Khushnood. (Jan. 2018). *Facebook Wall Post Processed Data*. [Online]. Available: http://dx.doi.org/10.17632/4dwzvcdsv3.1

ASIF KHAN received the B.Sc. degree in physics (Hons.) and M.C.A. degree in master of computer science and application from Aligarh Muslim University, India, and the Ph.D. degree (Hons.) in computer science and technology, in 2016, which was awarded by the University of Electronic science and Technology of China (UESTC's) Academic Achievement Award and Excellent Performance Award (2015–2016). He was an Adjunct Faculty at the University of Bridgeport, Bridgeport, CT, USA, for China Program in Summer 2016. Previously, he was a Visiting Scholar for Big Data Mining and Application at the Chongqing Institute of Green and Intelligent Technology (CIGIT), Chinese Academy of Sciences, Chongqing, China. He is currently a Postdoctoral Scientific Research Fellow at UESTC. He is also holding a position of Assistant Professor with BSA Crescent University, India. He is a contributor to many international journals with robotics & vision analyses about the contemporary world in his articles. His main interests are the machine learning, robotics vision, and new ideas regarding vision based information critical theoretical research.

JIANPING LI received the M.S. degree in computing mathematics and the M.E. degree in soft engineering from Xi'an Jiaotong University, in 1989, and the Ph.D. degree in computer science from Chongqing University, in 1998. As a visitor scholar, he has visited some famous universities around the world, from 1999 to 2006. He is the author and coauthor of 18 books on subjects ranging from wavelet analysis and its applications to computer science. He has published more than 200 technical articles. His current interests include wavelet theory and applications, fractals, image processing, pattern recognition, electronic commerce, and information security. He is the General Chairman of the First Conference on Wavelet Analysis and its Applications to Signal Processing of China (2000), the Associate Chairman of the Second International Conference on Wavelet Analysis and its Applications in Hong Kong (Hong Kong Baptist University 2001), the Chairman of the International Computer Congress 2004 (ICC04), the Chairman of the Second International Conference on Active Media Technology (ICAMT04), the Chairman of the International Conference 2007 on Information Computing and Automation (ICICA07), and the Chairman of ICACIA08.

NAEEM AHMAD received the Ph.D. degree in computer science from Jamia Millia Islamia (A Central University), New Delhi, and the master's degree in computer science and applications from Aligarh Muslim University, Aligarh, India, in 2012 and 2017, respectively. He worked as an Assistant Professor with the School of Network Engineering, Jiangxi Ahead Software Vocational and Technical College, Nanchang, China. He is currently serving as an Assistant Professor with the Department of Computer Applications, Madanapalle Institute of Technology and Science, Madanapalle, India. His research interests include mobile ad-hoc networks, wireless sensor networks, and information centric networking.

SHUCHI SETHI received the Ph.D. degree in cloud security from Jamia Millia Islamia (A Central University), New Delhi, in 2017. She is actively involved in research in the domain of information security. Her interests include cloud computing and security, networking, cybersecurity, security in the IoT, and cryptography. She takes up freelance assignments and delivers lectures in her areas of research interests.

AMIN UL HAQ received the M.S. degree in computer science. He is currently pursuing the Ph.D. degree with the School of Computer Science and Engineering, UESTC, China. He has a vast academic, technical, and professional experience in Pakistan. He is currently a Lecturer with Agricultural University Peshawar, Pakistan. His research areas include machine learning, medical big data, the IoT, E-Health and telemedicine, and algorithms. He is the author of some research articles. He is also associated with wavelets active media technology and big data.

SAROSH H. PATEL received the B.E. degree (Hons.) in electrical engineering from Osmania University, India, and the M.S. degree in electrical engineering and technology management and the Ph.D. degree in computer science from the University of Bridgeport, Bridgeport, CT, USA. He is currently an Assistant Professor with the School of Engineering, University of Bridgeport. His research interests include manipulator prototyping, industrial control, modular morphological robots, and robotic swarms.

SABIT RAHIM received the M.S. degree from Hamdard University, Pakistan, and the Ph.D. degree from the University of Science and Technology, Beijing, China. He is currently working as an Assistant Professor with the Department of Computer Sciences, Karakoram International University, Gilgit-Baltistan, Pakistan. His research interests include machine learning, cloud computing, the IoTs and ICT in education in rural areas of developing countries. He is also a member of the provincial ICT Policy for Education of Gilgit-Baltistan.